Cardiac Arrhythmias Identification by Parallel CNNs and ECG Time-Frequency Representation

Jonathan R Torres¹, K De Los Ríos², Miguel A Padilla¹

¹ Institute of Applied Sciences and Technology, Universidad Nacional Autónoma de México, Mexico
² Institute of Physics, Universidad Nacional Autónoma de México, Mexico

Abstract

Heart abnormalities cause about 26% of the deaths of illnesses in the world. Developing computational tools to the electrocardiogram (ECG) interpretation plays a critical role in the clinical diagnosis of Cardiac arrhythmias (CAs). Aims: This study aimed to develop an automated abnormal pattern recognition method for clinical decision support capable of detecting between 27 possible CAs. Proposal: An improved deep learning (DL) model was employed by using raw-data and time-frequency representation (TFR) images. Methods: A vast set of ECG records were filtered and normalized. They were segmented and transformed into two sets of 2-D images. TFR images were obtained through Wavelet Synchrosqueezing (WS). The VGG-16 network was chosen, modifying the weights of the inner layers to adapt the model to the CAs detection task. A 10-fold cross-validation method was executed. Different training hyperparameters were tested to find the best model. Results: The model performed accurately identifying CAs, with an overall unofficial S-score of 0.766. This model had a high performance in detecting healthy subjects with an F1 score of 0.83.

1. Introduction

Heart abnormalities are the first cause of death for illnesses worldwide [1]. Cardiac arrhythmias (CAs) are the most frequent causes of them and contribute to approximately 15% to 20% of all deaths [2,3].

CAs occur when the electrical impulses that coordinate the heartbeats do not work correctly. As a result, the heart beats too fast, too slow, or irregularly [4]. The electrocardiogram (ECG) is a widely used clinical tool that displays the heart’s electrical activity. The standard 12-lead ECG is the most commonly used to diagnose cardiac abnormalities and other heart diseases associated with the cardiac rhythm [5]. The early detection of CAs and its treatment for sudden cardiac deaths (SCDs) prevention represents a significant opportunity to reduce mortality further [4]. However, manual interpretation of the ECG is slowly, requires training personnel with a high degree of technical knowledge, and suffers from subjectivity.

Physicians detect the action potentials of the signals from the ECG recordings. With this information, it is possible to perform a morphological analysis of the components of the action potential and determine whether or not the patient has a CA. This procedure analyses the absence of P wave, distances and morphologies irregular of QRS complexes, and irregularity of the segments. Recent results have shown that physicians have an accuracy rate of 75% in the detection procedure of some CAs [6].

Computational tools that using automatic detection and classification of CAs can assist physicians in the diagnosis of the ECGs recorded. Recently, there have been increasing numbers of research focused on 12-lead ECG classification through machine learning (ML) and deep learning (DL) algorithms. Theoretically, many of these algorithms have been accurate in the identification of CAs. However, the successful result of these tests is a consequence of the use of small and homogeneous databases. The Physionet/CinC 2020 challenge has provided a vast database for this purpose [7].

2. Material and methods

The CAs detection method proposed here consists of the following stages: ECG data pre-processing (noise removal and data segmentation) and CAs classification (Signals transformation and final classification). In the first stage, the wavelet transform (WT) method was applied to denoise the ECG signal. Then, the signals were segmented in equal periods of duration, taking into account the morphology of the signal. In the second stage, the Wavelet Synchrosqueezing (WS) method was applied to obtain the TFR images, and the images of the raw signals were obtained too. Finally, the arrangement of parallel convolutional neural networks (CNNs) based on VGG-16 was trained from scratch for the identification of 27 types of CAs.
2.1. Data set

Six public databases of 12-leads ECG records coming from four different sources were used [7]. The databases as a whole have a total of 111 identified CAs, of which just 27 general classes were chosen to be assessed [7]. The dataset was divided randomly concerning the patient into three sets, training, validation, and testing. The local test set is formed with 15% of the total data; the training set is comprised by 70% of the data, and the validation set is 15% of the data.

2.3. Signals denoising and segmentation

Signals were purged, and irrelevant information was discarded, such as signals with invalid data. A denoising process was performed to eliminate outside signals related to the sampling procedure. The WT method was used for this purpose with the Daubechies4 (db4) function since it allows decomposing the input signal into low and high-frequency components with adequate regularization [8].

Each file in each database contains the information corresponding to the 12-leads ECG. The first 6 s of each of the entire database signals were segmented in time intervals of 1.2 s. This interval size allows obtaining the relevant information around each peak, regardless of the type of CA that the patient presents, thus having the information of each cycle of beats.

2.3. Signals and TFR images

Both sets of images were constructed by taking the segments of each signal. First, the segments were plotted as time series and saved as 64 x 64 greyscale images. The second set was built through the transformation of the signals using the Wavelet Synchrosqueezing (WS) method to obtain the TF features. With these features, the TFR images were obtained (e.g., see Fig. 2 b, d, and f).

The WS method used for TFR is based on CWT [9]. In this transformation, concentrated high-resolution TF patterns are obtained, from which instantaneous frequency lines can be identified. The instantaneous frequency \( \omega(a, b) \) for any point \((a, b)\) of the original signal with \( W_{a,b} \neq 0 \) is given by:

\[
\omega_{a,b} = -i(W_{a,b})^{-1} \frac{\partial}{\partial b} W_{a,b}
\]

(1)

Where \(a, b\) and \( W_{a,b} \) are the scale factor, translational value, and WT, respectively. From this instantaneous frequency, the Synchrosqueezing discrete transform is determined at a local frequency point given by the transformation [9]:

\[
T_{\omega_k,b} = (\Delta \omega_{a,b})^{-1} \sum_{a,k:|\omega_{a,k,b}\omega_k|<\Delta \omega/2} W_{a,k,b} a_k^{-2}(\Delta a)_k
\]

(2)

Finally, the representation of the signal in the TF space at high resolution is obtained. Figure 2 shows both the raw signal and the TFR of three different types of CAs with the same window length of 1.2 s.

Different features can be observed for each segment, especially the amplitude in the TFR and the morphology in the raw signal images. The representation of the signal segments in these two types of images gives relevant information that is no longer observed when using only one of the two types of representation shown here. Then, it is possible to observe notable differences both in each class and to each kind of representation.

Figure 1. Transformation of the 1.2 s segments extracted from the signals to images of raw signals and TFR image respectively a, d) segment of normal sinus rhythm (NSR), b, d) segment St depression (STD), and c, f) Premature atrial contraction (PAC).

As a result, we obtained 120 images for each patient from the arrhythmia databases used here. Figure 1 describes just six CAs images used as input of the classification stage.

2.4. Deep learning model

Convolutional neural networks (CNNs) have proven useful for automatic feature extraction in the detection of abnormal patterns in clinical images without the use of preprocessing algorithms or manual intervention [10].

A CNN is composed of an input and an output layer, as well as a large number of hidden layers that are commonly composed of convolutional, pooling, and fully connected layers. The convolutional layers are locally connected to extract the features by applying a set of weights called kernels. The ReLU function for an input value \( x \) is generally used as activation functions and is defined as:
\[
f(x) = \begin{cases} 
0, & \text{if } x < 0 \\
1, & \text{if } x \geq 0
\end{cases}
\]  

(3)

Relevant high-level features can be extracted with an increasing number of convolutional layers. The weights of the convolutional kernel parameters in each layer are trained with the backpropagation (BP) algorithm [11].

### 2.4.1. Model description

This deep network model provides the automatic classification of input segments through an end-to-end structure without the need for any hand-made feature extraction or selection steps.

The structure of the deep network is composed of an arrangement based on the VGG-16 network [11], where the feature extraction stages (convolutions) are duplicated and arranged in parallel. This allows entry of the two independently constructed image sets. At the end of this arrangement, there is a fully connected network that takes the output of convolution/pooling and predicts the best label to describe the image. The designed parallel CNNs model is shown in Fig. 2.

#### 2.4.1. Model architecture

The network input size parameters were modified to support the two sets of 64 x 64 grayscale images used as input of the model modified here.

During the training, the 120 images constructed per patient of each of the classes pass through two 64 x 64 x 64 convolutional layers in the input layer and later through a max-pooling layer. The information is subsequently transferred to the following general layers. The second layer consists of two convolutional layers and one of max pooling, and the next general layers are composed of three convolutional layers (see model characteristics in Fig. 2). At the end of the convolution layers of each available layer, there is a max-pooling layer. In convolutional layers are used filters with a kernel size of 3 x 3 and in the max-pooling layers a kernel size of 2 x 2.

After the five general layers of parallel CNNs, there is an FC of three layers with a different architecture. The first layer of FC receives the 4096 features obtained. The last layer, called a soft-max layer, contains 27 channels, and it is in charge of classifying 27 labels (one for each class, see network configuration in figure 3). Finally, the network gives the outputs to identify a CA.

### 2.5. Model evaluation

The proposed method was applied to the set of databases obtained. From the results of the classification with the designed model, some parameters were calculated: the F1-Score and average loss. To evaluate the effectiveness and reliability of the proposed parallel model for the CAs identification, S-score challenge was also computed [7]. Although the database size is large, it is necessary to carry out the k-folds cross-validation technique to stabilize the performance of the statistical model.

### 3. Results and discussion

The results of the implementation of the proposed model are shown. Also, a comparison of evaluation metrics is made with different training parameters that were tested to find the best model.

The cross-validation method of 10-folds was carried out to assess the model. In the local testing process, an F1 score of 82.5% was reached with a loss of 0.0617 and an overall unofficial S-score of 76.56%.

#### 3.1. Model optimization

For the optimization of the CNN, model parameters such as learning rate and the batch size are taking into account. The step of model parameter optimization is indispensable to achieve the best classification performance.

Table 1. F1 score and average loss (A. Loss) when performing batch size variation.

<table>
<thead>
<tr>
<th>Batch size</th>
<th>L. Rate</th>
<th>F1-Score</th>
<th>A. Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>0.804</td>
<td></td>
<td>0.0645</td>
</tr>
<tr>
<td>512</td>
<td>0.825</td>
<td></td>
<td>0.0617</td>
</tr>
<tr>
<td>256</td>
<td>0.818</td>
<td></td>
<td>0.0706</td>
</tr>
<tr>
<td>128</td>
<td>0.807</td>
<td></td>
<td>0.0641</td>
</tr>
<tr>
<td>64</td>
<td>0.812</td>
<td></td>
<td>0.077</td>
</tr>
</tbody>
</table>

A set of variations in the batch size value were carried...
3.2. Comparison with other models

As an additional test, the classification of the CAs was performed by implementing two other widely-used CNNs to compare them with the proposed model.

Table 2. Comparison of the performances of three CAs classification algorithms.

<table>
<thead>
<tr>
<th>Model</th>
<th>F-1 score</th>
<th>A. Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG-19</td>
<td>0.813</td>
<td>0.075</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>0.764</td>
<td>0.102</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.825</td>
<td>0.062</td>
</tr>
</tbody>
</table>

From the results, it was observed that although the training time of the model proposed here is longer than that of ResNet-50 and that the VGG-19 network is more robust than the VGG-16, the evaluation of test signals gives a very accurate result with modified VGG-16 (see the comparison in Tab. 2).

4. Conclusion

In this paper, we proposed a useful CAs classification model using a parallel CNN with ECG images based on the VGG-16 network. As an input, two sets of 64 x 64 grayscale images were transformed from database ECG records. Over 37134 12-leads ECG records were processed, and near to 4456080 ECG beat images were obtained with 26 types of CAs and the normal rhythm. The optimized CNN model was designed with considering essential concepts such as K-fold cross-validation. As a result, our proposed scheme achieved 91.57% SP, 82.5% F1-score, and 76.6% S-score. Our ECG classification result indicates that the identification of arrhythmia with 2-D images and the VGG-16 model would be a practical approach to the detection of CAs from 12-leads ECG signals.

Acknowledgments

This project was developed with the support of DGAPA-PAPIIT TA100920. J.R. Torres and K. De Los Rios would like to thanks CONACYT-Mexico and UNAM-PAEP for the support received for their Ph.D. studies. This research was partially developed in the Mitzili-UNAM Supercomputer.

References


Address for correspondence:
Jonathan R Torres-Castillo & Miguel A Padilla-Castañeda
Instituto de Ciencias Aplicadas y Tecnología (ICAT), UNAM, Cto. Exterior, Cd. Universitaria, Mexico City, 04510, Mexico jonathanmrc@comunidad.unam.mx
miguel.padilla@icat.unam.mx